A Multi-Method Exploration of Crime Hot Spots

An intramural project of the:

Crime Mapping Research Center[1]
National Institute of Justice

Presented at *1998 Academy of Criminal Justice Sciences (ACJS) Annual Conference.*

Introduction by:

Eric Jefferis
NIJ/CMRC

Contributing authors:

Jeff Burka, Indus Corporation and U.S. DOJ Criminal Division
John Eck and Daniella Drapkin, Jeffrey Gersh, Baltimore/Washington HIDTA
Elizabeth Groff, Eric Jefferis, Nancy La Vigne, Cyndy Mamalian, Joseph Szakas
NIJ Crime Mapping Research Center
Robert Langworthy, University of Alaska - Anchorage
Ned Levine, Ned Levine and Associates
Lorraine Mazerolle and Theresa Conover, University of Cincinnati
Stan Openshaw and Ian Turton, University of Leeds
Ronni Surban, Charlotte-Mecklenberg PD
Doug Williamson, Hunter College and the CUNY

[1] Points of view are those of the authors and do not necessarily represent the view of the U. S. Department of Justice or the National Institute of Justice.

A Multi-Method Exploration of Crime Hot Spots

Introduction

Many criminal justice agencies are taking advantage of rapid technological advancements in computer hardware and software in their efforts to more effectively and efficiently identify unusual patterns of criminal activity within their jurisdictions. These unusual clusters (hereafter "hot spots") of crime across time and space can be identified and analyzed using various methods, from simple visual interpretation of point data to calculating standard deviational ellipses and animating raster map images.

Still, the preferred technique for most police departments is to simply plot crime incident locations on a base map (Eck, Gersh and Taylor 1997) and visually interpret the distribution (Mamalian and LaVigne 1999). Several problems arise, however, in trying to visually interpret point maps. For example, repeated incidents may be represented by overlapping points, and closely situated points can appear cluttered and make the map uninterpretable. In addition, research has indicated that visual interpretations of point clusters vary among map-readers, and that cluster perceptions are affected by overlapping symbols (Sadahiro 1997). In other words, visual interpretations of point maps may be inconsistent.

Problems associated with visual interpretation of hot spots have led to increased reliance on more sophisticated spatial analysis methods. A recent survey indicated that computer aided hot spot identification techniques are being used by law enforcement agencies throughout the country (Mamalian and LaVigne 1999). The majority of departments that conduct crime cluster or hot spot analyses (86 percent) report using visual identification of hot spots, followed by use of a computer program that identifies hot spots (25 percent) (ibid.). Questions remain, though, about the appropriateness and validity of both visual and computer aided hot spot identification techniques. Each is thought to have its individual strengths and weaknesses, but the field lacks a systematic comparison of the techniques to verify these assumptions. In response to this need, the National Institute of Justice's (NIJ) Crime Mapping Research Center (CMRC) organized a systematic comparison of twelve hot spot identification techniques. In general, this comparison project sought to explore the methods' underlying processes as well as their utility and validity. The objective of the current paper is to briefly summarize the findings from that "Multi-Method Exploration of Crime Hot Spots" intramural research project. Links to the full evaluation reports are provided.

Underlying Issues in Hot Spot Analysis

Techniques to analyze point distributions, and to identify patterns within those points, began with the works of ecologists and botanists approximately sixty years ago (Chakravorty 1995). It has only been in recent years that criminal justice practitioners and researchers have realized the practical and theoretical value of identifying unusual crime patterns. For example, unusual crime clustering may prompt a crime analyst to recommend saturation patrol or some other response to that area. Theoreticians may explore the underlying social conditions that permit a hot spot of drug offending to flourish. It is critical to understand and consider the underlying assumptions of common hot spot analysis procedures before deploying resources or testing hypotheses so that we might avoid inappropriately allocating scarce manpower or misspecifying theoretical models. Therefore, before proceeding with the comparison of hot spot identification methods, we briefly address several issues that underlie hot spot analysis. (Refer to the referenced material for more detailed discussion of these matters.)

Foremost, readers should keep in mind that the null hypothesis we commonly test in hot spot analyses is that of no clustering (or Complete Spatial Randomness (CSR)) (Chakravorty 1995). In other words, each location in a study area is assumed to have an equal chance of having a crime point. Immediately, we recognize that this assumption will be violated when examining the distribution of crime incidents. The urban (or suburban or rural) environment does not distribute motivated offenders or suitable targets uniformly. Quite simply, people are not evenly distributed and are contained by natural and manmade boundaries. At the city level, high-rises are more densely populated than single family homes, and inner city neighborhoods are more densely populated than suburban neighborhoods. The assumption of CSR is further confounded by sparse offender populations on lakes, rivers and other uninhabitable geographic boundaries. More to the point, grid cell based analyses inappropriately assume homogeneity within each cell (ibid.). A grid overlay of any city will reveal that the populations at risk (e.g., persons, households, automobiles) are not evenly distributed to each of the cells and this often leads to the classic "hot spots" of crime in the central city. For this reason, Chakravorty (1995) makes a convincing argument that census geography polygons (i.e., tracts, block groups, blocks) might instead be the more appropriate base map with which to scan for hot spots. Finally, it should be noted that cluster analysis techniques such at the k-means discussed below also fail to control for the heterogeneous dispersion of populations at risk.

A second common hot spot identification assumption is that events are independent of one another, i.e., the existence of point A existing does not affect the likelihood of point B. In fact, though, the "first law of geography" stipulates that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1979). This brings us to the concept of spatial autocorrelation. Spatial autocorrelation is the result of like incidents tending to cluster because they are similarly influenced by similar processes (Chakravorty 1995). More formally, spatial autocorrelation can be defined as, "the coincidence of value similarity with locational similarity." (Anselin and Bera, 1996)[1]. For the purposes of the present comparison study, measures of local spatial autocorrelation provide an indication of unusual clustering (a hot spot).

The hot spot identification methods discussed in this paper can all be characterized as variations of five general techniques: (1) visual interpretation; (2) choropleth mapping; (3) grid cell analysis; (4) point pattern analysis, and (5) spatial autocorrelation as discussed above. Though the difficulties with visual interpretation of point clusters have been outlined above, it is important that they be considered in further detail. Overlapping and closely situated point distributions have been shown to complicate visual identifications of clusters and visual interpretations of point clusters vary among map-readers (Sadahiro 1997). Experienced crime analysts are certain to comment that efficient pin mapping relies more on using smaller spatial

[1] This notion of spatial autocorrelation is particularly important for researchers performing multivariate modeling and might suggest the need for inverse distance weightings.

and/or temporal subsets of the available data (Sorban 1998). For example, a week of street robbery data would be analyzed spatially rather than a month of street robberies. However, while this may help reduce the confusion created by cluttered and overlapping incidents, it does little to resolve the problem of different map-readers identifying different clusters.

Choropleth maps (a.k.a. graduated color maps) simply involve varying the shade of map unit polygons in response to different attribute values, e.g., homicide rates divided into quartiles with the highest quartile being shaded red, the third orange, the second yellow, and the fourth green. These types of maps provide a visual summary of data points (or rates) that are easy to differentiate and are supported by most mapping packages. Another key benefit of choropleth maps is that they remove assumptions about point accuracy; points geocoded to street centerline files provide only approximate locations on the street segment. At the same time choropleth maps include varying size polygons that can misleading the reader and, like grid cell analysis, in using choropleth mapping we cannot assume homogenous underlying at-risk populations.

Grid cell analysis is the analytic basis of the majority of the methods reviewed here. In its simplest form, grid cell analysis involves a grid of equal size square cells that is draped over the point coordinates. Points are then aggregated to the cell in which they fall *or* within a specified radius from the cell center. Further analyses for unusual clustering are then performed on the aggregated grid cell counts. When performing grid cell analyses, the selected cell size is critical. While smaller cells will have greater resolution and accuracy they will require greater computing power.

Background

In September of 1997, the CMRC hosted a meeting of experts in the field of crime mapping to begin a dialogue on the issue of crime hot spot analysis and to establish a means to continue that discussion[2]. Specifically, the meeting was convened to: (1) establish outstanding questions regarding hot spot analysis; (2) begin to identify limitations of existing hot spot identification methods; and (3) brainstorm about possible new methods. Based upon the discussion of these issues, the participants felt that an in-depth look at available techniques was a necessary first step. As such, several of the participants agreed to participate in a research project to systematically compare twelve hot spot identification techniques, many of which are currently used by law enforcement and the research community. CMRC staff solicited several additional researchers for the project and organized the research activity.

Method

A convenience sample of twelve hot spot identification methods was chosen for evaluation. This sample was selected from a list developed by CMRC staff of hot spot identification methods they knew were being used in law enforcement or with which they were otherwise familiar. A group of researchers familiar with crime mapping and GIS applications in general were solicited to evaluate the twelve methods. With only two exceptions, the researchers were chosen for their *lack* of familiarity with the method, because one goal of the project was to

[2] The meeting proceedings are available from the editor upon request.

evaluate the "user friendliness" of the method; we would expect expert users, or the developer, to have difficulty providing an objective evaluation of this criterion[3].

At this point we would like to add a word of caution about the interpretation of these findings. As with any research endeavor that involves multiple data collectors, there exists the possible problem of inter-rater reliability. In other words, the subjective evaluation criteria laid out below leave significant room for evaluator latitude; we caution the reader to avoid making summary judgments based upon this work alone. Rather, we suggest that these evaluation reports serve as the starting point to better understanding the application and utility of hot spot identification methods. Additionally, we refer the readers to the web site addresses of the software developers listed in Table 1.

Table 1. Hot Spot Identification Method and Contact Information

HOT SPOT IDENTIFICATION METHOD	CONTACT FOR ADDITIONAL INFORMATION
Geographic Analysis Machine (GAM)	www.ccg.leeds.ac.uk/smart/gam/gamin.html
SaTScan	www.dcpc.nci.nih.gov/bb/Software.html
Spatial and Temporal Analysis of Crime (STAC)	www.acsp.uic.edu
Vertical Mapper	www.northwoodgeo.com
Spatial Analyst	www.esri.com
IDRISI	www.idrisi.clarku.edu
K-means	**n.a.**
CrimeStat Kernel Density Interpolation	www.nedlevine.com
SpaceStat	www.rri.wvu.edu
"Repeat Place" Analysis	**n.a.**
Visual Interpretation	**n.a.**

Note: n.a. indicates a specific contact site is not available.

Data for this project included all burglaries and street robberies[4] in Baltimore County from November 1, 1996 through November 30, 1997. A total of 96.5 percent of the addresses were successfully geocoded (assigned an *x* and *y* coordinate) producing a total sample size of 7,719 cases. These data were provided by the Baltimore County Police Department[5] and are routinely used as part of their Regional Crime Analysis System (RCAS).

In order to minimize the subjectivity of the evaluation, each researcher was assigned three tasks with loosely restrictive parameters. Each was asked to produce four hot spot maps, provide a written evaluation of the software package used, and present their findings at an Academy of Criminal Justice Sciences Annual Meeting workshop (1998). The four hot spot maps included: (1) residential burglary at the county level; (2) residential burglary in southwest Baltimore; (3) street robbery at the county level; and (4) street robbery in southwest Baltimore.

---

[3] Two exceptions were made to this selection criterion. The Geographic Analysis Machine (GAM) software was evaluated in part by its developer and an expert user evaluated the Spatial Analyst software.

[4] These two offenses were chosen to provide significant variation in the frequency of reported offenses (6,054 burglaries and 1,188 robberies were reported)

[5] We extend our appreciation to Phil Canter of the Baltimore County Police Department for providing the data for this project.

In their written evaluation of the method, each researcher was instructed to answer the following questions:

1. **What is the underlying algorithm?** An important first step in understanding each hot spot program is to understand its underlying algorithm or process by which hot spots are identified. If this information is not accessible or available, that is also important. For the subjective mapper who was not using a computer or statistical program, we wanted to understand the cognitive process in hot spot identification (e.g., Did she look at bordering neighborhoods? How did she determine what was non-random clustering?).
2. **How user friendly is the program?** Are data imports/exports complicated? Can the user import/export a variety of file types? Can the user set or adjust parameters? Does map production require exporting data to another program?
3. **Do the resulting hot spots have face validity?** This is largely a subjective assessment of whether the results the user obtains appear to fit the data.
4. **Does the program have practical utility?** Can the program be integrated with other programs? Is it better suited for practitioners or researchers? Are resulting hot spots conducive to statistical analyses? Can changes in hot spots be assessed? How well does the program handle secondary data sources? Is the program more suitable for common or rare events?
5. **Is the program flexible?** Do you have the ability to make changes to parameter selections, or is this a cumbersome process? Can underlying population estimates be included as a denominator?
6. **What did you like best and least about this product?** If the researcher had the opportunity to talk with the developer, what recommendations would he make?

Findings

Links to the complete evaluation reports are provided below:

Geographic Analysis Machine (GAM)
SaTScan
Spatial and Temporal Analysis of Crime (STAC)
Vertical Mapper
Spatial Analyst
IDRISI
K-means
CrimeStat
Kernel Density Interpolation
SpaceStat
“Repeat Place” Analysis
Visual Interpretation

Conclusions

The general hot spot issues and summary results presented here only begin to illustrate the intricacies of hot spot analysis. We had few assumptions going into this comparative analysis about how much agreement would be found among the methods. Obviously, we expected variations based on the parameters that could not be uniformly set across methods, but the magnitude of the differences found between methods was not anticipated. While it was not possible to provide strong, objective measures of all the evaluation criteria, the results indicate a need for more practical and user friendly methods.

It is clear that much work is left to be done in this area. We hope that this project will serve as a building block of what is known about hot spot analysis, and that it can be developed further as new methods are employed and old methods are updated.

REFERENCES

Anselin, L. and A. Bera (August 1996) "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." #96-0128.

Chakravorty, S. (1995) "Identifying Crime Clusters: The Spatial Principles." Middle States Geographer, 28.

CMRC. (Forthcoming) "The National Crime Mapping Survey." Report prepared for the National Institute of Justice, Crime Mapping Research Center.

Eck, J., Gersh, J., and C. Taylor (1997). "Mapping Hotspots of Crime and Related Events." Paper presented at the conference on the spatial analysis of crime sponsored by the New York City Police Department, the Center for the Applied Studies of the Environment, and the Center for Urban Research at CUNY.

Hjalmars, U., Kulldorff, M., Gustafsson, G., and N. Nagarwalla (1996) "Childhood Leukemaemia in Sweden: Using GIS and a Spatial Scan Statistic for Cluster Detection." Statistics in Medicine, 15:707-715.

Kulldorff, M. (1997) "A Spatial Scan Statistic." Communications in Statistics: Theory and Methods, 26:1481-1496.

Kulldorff, M. and N. Nagarwalla (1995) "Spatial Disease Clusters: Detection and Inference." Statistics in Medicine, 14:799-810.

Kulldorff, M., Feuer, E., Miller, B., and L. Freedman (1997) "Breast Cancer in Northeastern United States: A Geographical Analysis." American Journal of Epidemiology, 146:161-170.

Mamalian, C., and N. La Vigne (1999) NIJ Research Preview: The Use of Computerized Crime Mapping by Law Enforcement: Survey Results. Washington, DC: National Institute of Justice, U.S. Department of Justice.

Sadahiro, Y. (1997) "Cluster Perception in the Distribution of Point Objects." Cartographica, 34(1).

Tobler, W. (1979) "Cellular Geography." In Philosophy in Geography, Ed. S. Gale and G. Olson. Pp. 379-386. Dordrecht: Reidel.